

# Estimating the p-mode frequencies of the solar twin 18 Sco\*

M. Bazot<sup>1</sup>, T. L. Campante<sup>1</sup>, W. J. Chaplin<sup>2</sup>, H. Carfantan<sup>3</sup>, T. R. Bedding<sup>4</sup>, X. Dumusque<sup>1,5</sup>, A.-M. Broomhall<sup>2</sup>, P. Petit<sup>3</sup>, S. Théado<sup>3</sup>, V. Van Grootel<sup>3</sup>, T. Arentoft<sup>6</sup>, M. Castro<sup>7</sup>, J. Christensen-Dalsgaard<sup>8</sup>, José-Dias do Nascimento Jr.<sup>7</sup>, B. Dintrans<sup>3</sup>, H. Kjeldsen<sup>8</sup>, M. J. P. F. G. Monteiro<sup>1,9</sup>, N. C. Santos<sup>1,9</sup>, S. Sousa<sup>1</sup>, and S. Vauclair<sup>3</sup>

<sup>1</sup> Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762, Porto, Portugal; bazot@astro.up.pt

<sup>2</sup> School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, United-Kingdom

<sup>3</sup> Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse UPS-OMP / CNRS, 14, avenue Édouard Belin, 31400 Toulouse, France

<sup>4</sup> Sydney Institute for Astronomy, School of Physics, University of Sydney NSW 2006, Australia

<sup>5</sup> Observatoire de Genève, 51 Chemin des Maillettes, CH-1290, Sauverny, Suisse

<sup>6</sup> VUC Aarhus, Ingerslevs Boulevard 3, 8100 Aarhus C., Denmark

<sup>7</sup> Depart. de Física Teórica e Experimental, Univers. Federal do Rio Grande do Norte, 59072-970 Natal, R.N., Brazil

<sup>8</sup> Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

<sup>9</sup> Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

Preprint online version: September 4, 2012

## ABSTRACT

Solar twins have been a focus of attention for more than a decade, because their structure is extremely close to that of the Sun. Today, thanks to high-precision spectrometers, it is possible to use asteroseismology to probe their interiors. Our goal is to use time series obtained from the HARPS spectrometer to extract the oscillation frequencies of 18 Sco, the brightest solar twin. We used the tools of spectral analysis to estimate these quantities. We estimate 52 frequencies using an MCMC algorithm. After examination of their probability densities and comparison with results from direct MAP optimization, we obtain a minimal set of 21 reliable modes. The identification of each pulsation mode is straightforwardly accomplished by comparing to the well-established solar pulsation modes. We also derived some basic seismic indicators using these values. These results offer a good basis to start a detailed seismic analysis of 18 Sco using stellar models.

**Key words.** Stars: individual: 18 Sco - Stars: oscillations - Techniques: radial velocities - Methods: data analysis

## 1. Introduction

In the field of stellar physics, the study of solar twins has recently received growing attention. Since the term was first coined by Cayrel de Strobel et al. (1981) to designate stars spectroscopically identical to the Sun, they have been the focus of photometric and spectroscopic studies aiming at measuring their fundamental atmospheric parameters (Gustafsson 1998; Meléndez & Ramírez 2007; Gustafsson 2008; Meléndez et al. 2010). On the one hand there has been an on-going race to find the “best” solar twin. On the other, samples of such stars have been used to try to answer statistically the question: is the Sun a peculiar star?

Nowadays, thanks to various technical breakthroughs in the field of observational astrophysics, these founding studies can be supplemented by additional measurements. Spectropolarimetric surveys of the solar twins using NARVAL have allowed observers to detect magnetic fields in several objects and to reconstruct their magnetic topology (Petit et al. 2008). More recently, Bazot et al. (2011, hereafter Paper I) combined interferometric and asteroseismic measurements, from respectively the PAVO beam combiner at the CHARA array and the high-precision spectrograph HARPS at La Silla Observatory, to estimate the linear and acoustic radii of 18 Sco, the brightest solar

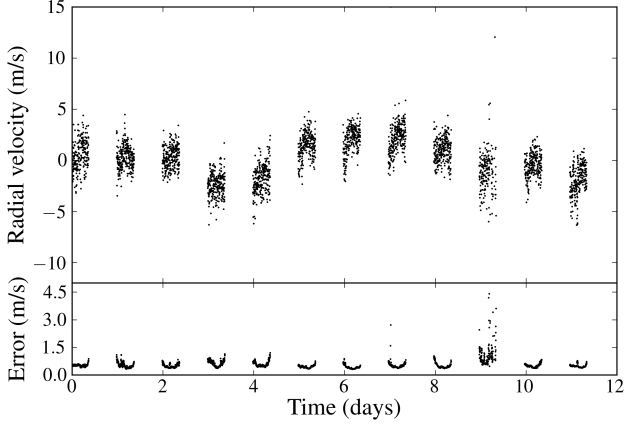
twin, and to derive its mass. The method used to determine the acoustic radius relied on the use of the autocorrelation of the radial-velocity time series. This paper is the continuation of this study, which now aims at a detailed analysis of the seismic data.

Asteroseismology measures quantities directly sensitive to the stellar interiors. This is an advantage compared to the classical observable quantities, usually obtained from spectroscopy or photometry, which are sensitive to the poorly-modelled external layers of the stars. In contrast, it is possible to extract information from the seismic signal roughly independent of these layers and more robustly described by the existing stellar codes.

Following the development of high-precision spectrographs (mostly for the search of extrasolar planets) and the space missions CoRoT and Kepler, seismic data have been more and more frequently used to model stars in general, and main sequence sun-like stars in particular (e.g., Miglio & Montalbán 2005; Bazot et al. 2005, 2008; Doğan et al. 2010; Metcalfe et al. 2010; Brandão et al. 2011). The best-case scenario from the perspective of detailed modelling is to have access to the eigenfrequencies of the stellar pulsation modes and to their characteristic numbers  $n, l, m$ <sup>1</sup>. They are then combined to effectively obtain

\*Based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere, Chile (run ID: 183.D-0729(A))

<sup>1</sup>With  $n$  the number of nodes of the eigenfunction to the stellar pulsation equations and  $l, m$  corresponding to the angular degree and the azimuthal order of the eigenfunction, related to the orthogonal set of spherical harmonics  $Y_l^m$ , which are solutions to Laplace’s equation.



**Fig. 1.** Time series of radial velocities (upper panel) and their uncertainties (lower panel) from HARPS observations of 18 Sco. A few points deviating strongly from the bulk of the time series lie outside the plotted range.

surface-independent information. In the following, we present the strategy used to determine these frequencies.

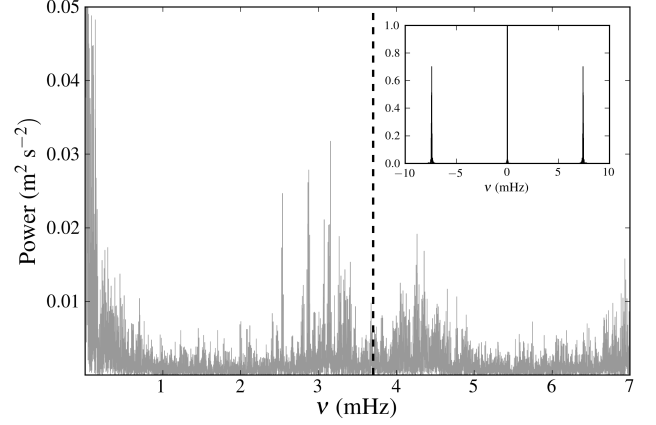
In Section 2 we discuss the data and return to some of the sampling issues evoked in Paper I. Before proceeding to the frequency analysis we give an overview of the characteristics of the noise affecting the seismic signal in Section 3. We recall the characteristics of the parametric model used to describe the power spectrum in Section 4. We define an inverse problem for the estimation of the parameters and cast it into a Bayesian formulation. We then solve it numerically using a Markov Chain Monte Carlo (MCMC) algorithm. Our strategy is to first, test the methodology using simulated time-series and then to apply it to the real data. In Section 5, we discuss several aspects of our results. From a methodological standpoint, we try to assess the robustness of our MCMC strategy by comparing it to another estimation procedure. We also discuss the choice of our parametric model and the priors we used in our Bayesian formulation. From a physical point of view, we measure the impact of our new estimates on the acoustic radius and the stellar mass.

## 2. Data

The time series was described in Paper I and is shown in Fig. 1 in its unfiltered version. We collected 2833 radial velocity measurements over 12 nights from 10 to 21 May 2009. In this section, we would like to draw the attention on some of the consequences of the window function  $w(t) = \sum \delta(t - t_n)$ , with  $t_n$  the median time of the  $n$ -th exposure.

The observed signal can be written  $y(t) = \tilde{y}(t)w(t)$ , with  $\tilde{y}(t)$  the continuous time series. When the sampling is uniform, the definition of the Nyquist frequency is  $\nu_N = 1/(2\Delta t)$ , with  $\Delta t$  the sampling time. In the case of unevenly sampled time series, there is, strictly speaking, no Nyquist frequency. However, this does not mean that there is no spectral folding and equivalent frequencies have been suggested (see e.g. [Bretthorst 2000](#)).

The power spectrum of 18 Sco is shown in Fig. 2. The differences with the figure of Paper I come from the filtering of the time series<sup>2</sup>. The inset shows the spectral window, that is the squared modulus of the Fourier transform of  $w(t)$ . Its maximum



**Fig. 2.** Power spectrum of the radial velocity of 18 Sco estimated using a Lomb-Scargle weighted periodogram. The vertical dashed line marks the equivalent Nyquist frequency calculated using the median exposure time. The inset shows the power spectrum of the window function  $w(t)$ , normalized to its maximum.

is at 0 (around which also stand the daily aliases, not visible here, see Paper I), and two prominent peaks at  $\pm 7.4$  mHz. They are the aliases caused by the sampling. This means that any signal present at frequencies around 3.7 mHz is likely to be folded. It should, however, be emphasized that this is not a clear-cut limit, nor an exact estimate; therefore, it is not possible to assess firmly whether some modes at higher frequencies perturb the spectrum below 3.7 Hz or, for that matter, if we are missing some genuine modes above this limit.

It turns out that the median observing time,  $\text{med}(\Delta t)$ , offers a reasonably good estimator of the upper limit for folding. This is expected because the sampling is close to uniform. From night to night, the mean values of the observing time lie in the range 132 – 172 s and their standard deviations within 2 – 232 s (the upper value resulting mostly from one cloudy night, otherwise, the standard deviations lie in the range 2 – 60 s, the remaining discrepancy being explained by longer exposures during some of the nights). Furthermore, the median is almost unaffected by the daily gaps. Defining our “equivalent Nyquist frequency” as

$$\nu_N^{(n.u.)} = \frac{1}{2\text{med}(\Delta t)}, \quad (1)$$

we obtain an indicative value of 3.7 mHz, in good agreement with the general shape of the spectrum.

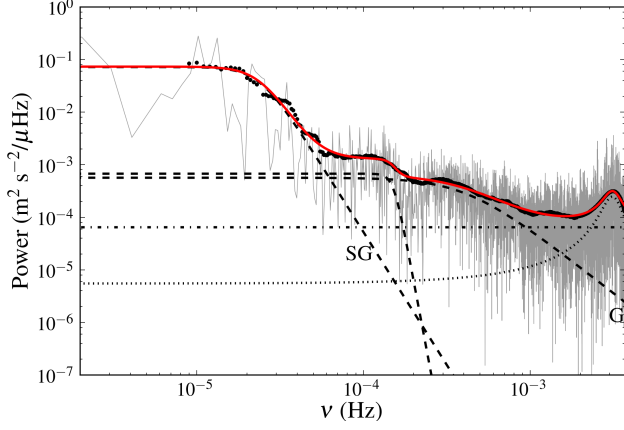
Such a low value of  $\nu_N^{(n.u.)}$  is also a problem when estimating the various noise contributions in the power spectrum. In particular, it makes it difficult to address the question of the photon noise level, which is the main source of noise at high frequencies, i.e. few mHz above the p-mode envelope (see Fig. 3). A common, straightforward approach consists of estimating it by simply calculating an average of the amplitude spectrum well above the p-mode region; this cannot be done here because of the spectral folding.

## 3. Noise

Before considering the extraction of p-mode oscillation frequencies, we discuss briefly the impact of the different noise sources

be considered as the reference. This does not affect the results from Paper I.

<sup>2</sup>And from a small numerical bug in the weighting scheme of the LS spectrum in Paper I. The lower amplitude observed in Fig 2 should



**Fig. 3.** Power spectrum of the radial velocity of 18 Sco represented in logarithmic units. The contributions to the noise from granulation (G), supergranulation (SG) and from a third Harvey-like component are represented as dashed lines. The photon noise is a dot-dashed line. The p mode component of the spectrum appears as a dotted line. The full red line shows their sum. The black dots are the points of a binned power spectrum that has been effectively used to perform the fit.

on the relevant signal. It is customary to consider three main phenomena contributing to the overall noise: activity, granulation (at different scales) and instrumental noise (which we assumed to be close to the photon noise, however see [Dumusque et al. 2011](#)). Characteristic times for activity typically scale as the rotational period. In the case of 18 Sco, a 22.7-d rotation period has been measured from spectropolarimetry ([Petit et al. 2008](#)). A 12-d time series is certainly not enough to capture this kind of signal. We therefore ignore this contribution.

[Harvey \(1985\)](#) suggested modelling the contributions to the power spectrum coming from granulation as Lorentzian,  $4\sigma^2\tau/[1+(2\pi\nu\tau)^2]$ , with  $\tau$  the characteristic timescale of the process and  $\sigma$  the velocity rms. Such functions are representative of non-oscillatory velocity fields whose autocorrelations decay exponentially with time. However, it is difficult to detect this kind of signal precisely enough to assess accurately which of these phenomena contribute effectively to the low-frequency signal. In fact, the exact dependency of these noise components on frequency is unclear and still subject to discussion (e.g. [Guenther et al. 2008](#)). Note that it is also possible to use simpler scaling laws of the form  $\nu^{-2}$  to model them. In all cases, following [Harvey \(1985\)](#), we should warn that these models are crude.

In seismic studies, it is customary to use parametrized “Harvey functions”,  $\mathcal{H}(\nu) = \alpha/[1 + (2\pi\nu\tau)^\beta]$ , in order to improve the fit to these low-frequency components. After several trials we found that we needed to use three such functions. The photon noise is assumed to be white. The contribution to the power spectrum of the p modes is modelled as a Lorentzian envelope ([Dumusque et al. 2011](#)). Figure 3 shows the power spectrum corresponding to our time series in logarithmic units. It has been computed using a weighted Lomb-Scargle periodogram ([Lomb 1976](#); [Scargle 1982](#); [Zechmeister & Kürster 2009](#)). It was evaluated at frequencies separated by  $1/T$ , with  $T$  the total observing time. To allow comparison with previous studies, it has been normalized according to [Kjeldsen et al. \(2005\)](#). The best fit to the spectrum of this composite model is also shown. For stability, the fit to the background was performed on a heavily smoothed version of the spectrum in order to retain only the slowly-varying

**Table 1.** Parameters for the best fit to the smoothed spectrum with uncertainties. SG and G stand respectively for the supergranulation and granulation components.

	$\tau$ (s)	$\alpha$ ( $\text{m}^2 \text{s}^{-2}/\mu\text{Hz}$ )	$\beta$
1 (SG)	$6642 \pm 438$	$(7.2 \pm 1.1) \times 10^{-2}$	$5.1 \pm 0.3$
2	$1051 \pm 34$	$(6.8 \pm 0.5) \times 10^{-2}$	$16.9 \pm 1.3$
3 (G)	$389 \pm 5$	$(5.7 \pm 0.1) \times 10^{-4}$	$2.47 \pm 0.05$
Power ( $\text{m}^2 \text{s}^{-2}/\mu\text{Hz}$ )			
Photon noise	$(6.46 \pm 0.16) \times 10^{-5}$		

features. These values, because of the low “Nyquist” frequency, should be considered with care. In particular, it is much easier to perform this fit when the photon-noise constant component can be fixed separately using the signal-free high-frequency regions of the power spectrum<sup>3</sup>.

The results of our fit are given in Table 1, which displays the parameters of the Harvey functions and the average photon noise. In order to estimate uncertainties on the fitted parameters, we assumed that the binned points in the smoothed power spectrum obey Gaussian statistics, whose variances were evaluated using the rms in each bin. We then used a simple Monte Carlo experiment with 10 000 simulated spectra to obtain the uncertainties displayed in Table 1.

The Harvey functions are often interpreted as representing different scales of surface convective motions. In the case of granulation, the time scale agrees well with those found in the solar case by [Harvey \(1985\)](#) and [Lefebvre et al. \(2008\)](#). It is possible that the larger time-scale phenomena correspond to supergranulation, which has been well-studied in the solar case. We however note that the value returned for  $\tau$  is lower by one or two orders of magnitude than results found in the by [Harvey \(1985\)](#); [Title et al. \(1989\)](#); [De Rosa et al. \(2000\)](#) and [Shine et al. \(2000\)](#). It is not possible with such short time series to determine whether it is an intrinsic difference between 18 Sco and the Sun or if this is a methodology-induced effect. The intermediate scale is sometimes interpreted as mesogranulation ([Dumusque et al. 2011](#)), but its very existence is subject to debate in the solar case (see [Nordlund et al. 2009](#), and references therein). This debate being outside the scope of this paper, we limit ourselves to mention this intermediate scale without attributing it to surface convection or to an instrumental artifact in the data.

## 4. Modelling of the signal

### 4.1. The spectrum model

In sun-like stars, p modes are excited by turbulent convective motions. Meanwhile, each mode is damped (see e.g. [Houdek et al. 1999](#)). It is assumed that the number of excitation events per damping time is large. The central-limit theorem then ensures that the amplitude distributions of the modes converge towards normal ones ([Foglizzo 1998](#)), which translates into a power spectrum that is exponentially distributed. We make the additional assumption that the frequency bins in the Fourier space are independent. In this case, the density probability density of the power spectrum at frequency  $\nu_i$  is given by

<sup>3</sup>In fact, the only region of the power spectrum that seems relatively free from stellar granulation noise is the narrow range 1800–2200  $\mu\text{Hz}$ , where the average noise is 2.6  $\text{cm s}^{-2}$  (against 2.4  $\text{cm s}^{-2}$  for the fitted photon noise).



$$f(p(v_i)) = \frac{1}{\mathcal{P}(v_i)} \exp \left[ -\frac{p(v_i)}{\mathcal{P}(v_i)} \right]. \quad (2)$$

Ideally, if one obtains many independent measurements of the stellar power spectrum, it will tend, at each  $v_i$ , towards the expectation value  $\mathcal{P}(v_i)$ . The form of this function must be specified from our knowledge of the physical processes governing the mode excitation and the various sources of noise. The expected value of the power spectrum corresponding to one oscillation mode  $k$ ,  $\mathcal{P}_k(v_i)$ , was derived theoretically by [Anderson et al. \(1990\)](#). Considering the equation for a simple damped oscillator with random forcing, they showed that  $\mathcal{P}_k(v_i)$  has the form of a Lorentzian centred at  $v_k$ , the eigenfrequency of the mode. If we make the additional assumption that the modes are uncorrelated, which is supported by the observations in the solar case ([Foglizzo et al. 1998](#)), the expectation value for the multi-modal power spectrum can be written in the form

$$\mathcal{P}(v_i) \simeq \left[ \mathcal{P}'(v_i) + \mathcal{B}(v_i) + \sum_{k=1}^K \frac{V_k^2 H_k}{1 + u_{i,k}^2} \right] * |W(v)|^2. \quad (3)$$

Here, the  $V_k$  represent the visibilities of the modes,  $H_k$  their heights, and  $u_{i,k} = 2(v_i - v_k)/\Gamma_k$ , with  $\Gamma_k$  the linewidth of the Lorentzian. It is related to the mode lifetime through the relation  $\Gamma_k = 1/\pi\tau_k$ , with  $\tau_k$  the lifetime of mode  $k$ . These parameters are real-valued.  $\mathcal{P}'(v_i)$  is the power density for all the modes that are not taken into account in the sum (mostly  $l = 4$  and  $5$  modes, but also unidentified modes with  $l \leq 3$ ; [Fletcher et al. 2009](#)) but still contribute to the overall power, and  $\mathcal{B}(v_i)$  the noise contribution, using a  $\nu^{-2}$  scaling law to describe the low-frequency component. Note that the convolution by  $|W(v)|^2$  ( $W(v)$  being the Fourier transform of the window function) in Eq. (3) is only justified because we are considering the expectation value of the power of a stochastic function ([Deeming 1975](#)), hence assuming our time series are purely stochastic functions. A more rigorous approach can be found in [Stahn & Gizon \(2008\)](#) but was not implemented here.

The model is characterized by a parameter vector  $\theta \in \{\{\theta_k\}_{1 \leq k \leq K}, \theta_{\mathcal{B}}, \theta_{\mathcal{P}'}\}$ , which we must estimate. Here  $\theta_k \in \{v_k, \Gamma_k, H_k\}$  and has to be estimated for  $K$  modes<sup>4</sup> (the  $V_k$  can be computed straightforwardly provided some assumptions are made on the mode degree). The vectors  $\theta_{\mathcal{B}}$  and  $\theta_{\mathcal{P}'}$  regroup the parameters describing respectively  $\mathcal{B}$  (see Sect. 3) and  $\mathcal{P}'$ . The inverse problem is now set, as we need to provide estimates of  $\theta$  being given  $\mathbf{y} = (y_1, \dots, y_N)$ . This estimation problem, when applied to helio- or asteroseismic data is sometimes referred to as “peak bagging”.

A classical approach, first used in helioseismology and then adopted in asteroseismology, involves estimating the parameters of  $\mathcal{P}(v)$  by minimizing the corresponding negative log-likelihood function ([Anderson et al. 1990](#))

$$-\ln(L(\theta)) = -\sum_i \ln f(p(v_i; \theta)) = \sum_i \left[ \frac{p(v_i)}{\mathcal{P}(v_i)} + \ln \mathcal{P}(v_i) \right], \quad (4)$$

where we used the notation  $f(p(v_i; \theta)) = f(p(v_i))$  to indicate the dependency on the model parameters<sup>5</sup>.

<sup>4</sup>This number has to be provided by some ansatz.

<sup>5</sup>Using the (frequentist) Maximum Likelihood (ML) method, one defines a Maximum Likelihood Estimator (MLE) as the value of  $\theta$  that minimizes  $-\ln(L(\theta))$

$$\text{MLE}(\theta) = \underset{\theta}{\operatorname{argmin}}(-\ln(L(\theta))).$$

#### 4.2. Bayesian formulation of the problem

Studying  $L(\theta)$  can be an challenging task. For likelihoods that are highly non-linear in the parameters, the chance is high that they possess multiple maxima. This makes the identification of high-likelihood regions equivocal and hence complicates the parameter estimation. Furthermore, if the dimension of the space of parameters is large, serious algorithmic problems might occur. In particular, it becomes difficult to even locate the maxima of  $L(\theta)$  because one cannot sample efficiently the space of parameters.

A possible approach to this problem consists of using the Bayes' formula

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta)L(\theta), \quad (5)$$

where  $\pi(\theta|\mathbf{y})$  is the Posterior Probability Density (PPD) of the parameters for fixed data  $\mathbf{y}$  and  $\pi(\theta)$  is the prior density on the parameters. We can therefore explicitly include the knowledge we already have on  $\theta$  in the PPD, so that  $\pi(\theta|\mathbf{y})$  becomes the quantity of interest instead of  $L(\theta)$ . If the prior is adequately chosen it might significantly restrain the volume to sample in the space of parameters. This potentially offers two advantages: (i) from the statistical point of view, by reducing the number of local maxima, (ii) numerically, some algorithms being more likely to converge when the volume in the parameter space decreases. For general reviews on the Bayesian methodology, we refer to monographs by [Gregory \(2005\)](#) or [Robert \(2007\)](#).

This approach has become very popular in the case of asteroseismology of sun-like stars. The main reason is that realistic priors on the frequency distribution of the pulsation modes can be obtained by using the theoretical asymptotic distribution of the p modes<sup>6</sup>.

Bayesian methodology applied to asteroseismology and how it might improve the fitting of the data has been discussed by, e.g., [Brewer et al. \(2007\)](#) and [Gaulme et al. \(2009\)](#). In the following, we will mostly focus on the specific setups used in our algorithms, with a particular emphasis on the priors included in our probabilistic models.

#### 4.3. Markov Chain Monte Carlo estimation

A widespread approach to Bayesian estimation consists in using Markov Chain Monte Carlo (MCMC) algorithms to approximate the PPD of the parameters.

The (stochastic) sampling in the parameter space relies on the convergence properties of Markov Chains which, under certain circumstances, generate realizations of a random variable according to a stationary distribution. The underlying idea of MCMC algorithms is thus to produce a Markov Chains whose stationary distribution is the target PPD ([Robert & Casella 1999](#)). MCMC algorithms are a class of numerical methods that aim at approximating PPDs. One of their features is that one can sample a complex distribution using much simpler ones (often called instrumental laws).

The base used here is a Metropolis-Hastings algorithm ([Metropolis 1953](#); [Hastings 1970](#)). Because the complexity of the sampling increases with the dimension of the parameter

<sup>6</sup> The asymptotic relation in the limit of high frequencies (high orders) and low degrees was given to successive orders of approximation by [Vandakurov \(1967\)](#) and [Tassoul \(1980\)](#) and can be written as  $\nu_{n,l} = (n + l/2 + \epsilon)\Delta\nu + O(\nu^{-1})$ , with  $\Delta\nu$  the average large separation, i.e. the inverse of twice the stellar acoustic radius, and  $\epsilon$  a phase-related constant.

space, an additional scheme was added in order to set the instrumental law used for the sampling. This was done in a burn-in sequence, during which the chain converges towards its stationary distribution<sup>7</sup>. The algorithm also allows one to run multiple parallel chains in order to use tempering, which is useful to prevent the Markov Chains becoming stuck in local minima. A complete description of the algorithm was given by Handberg & Campante (2011).

#### 4.3.1. Priors and algorithmic setup

The prior density on the frequency is certainly the most characteristic feature of asteroseismology of sun-like star. We first assume that the components  $\theta_{\mathcal{B}}$  and  $\theta_{\mathcal{D}}$  of  $\theta$  are independent from the other parameters. We fix them before the estimation of the rest of the parameters. This assumption might affect the estimation of the frequency of some low-amplitude mode. However, it improves significantly the convergence of our algorithm. We can reduce the prior probability density to

$$\pi(\theta) = \pi(\mathbf{v}, \Gamma, \mathbf{H}), \quad (6)$$

where we used the simplifying notation  $\mathbf{v} = (\nu_1, \dots, \nu_K)$ ,  $\Gamma = (\Gamma_1, \dots, \Gamma_K)$  and  $\mathbf{H} = (H_1, \dots, H_K)$ .

In practice, we recast the problem of estimating the  $K$  linewidths  $\Gamma_k$  to the estimation of only two parameters by assuming a linear dependency of the linewidth on the frequency. We considered only the values of the linewidth  $\Gamma_1$  and  $\Gamma_2$  at 2800  $\mu\text{Hz}$  and 3600  $\mu\text{Hz}$  respectively. From theoretical modelling and observational results in the Sun, it is clear that this is an oversimplification. However, it is difficult to suggest a proper empirical law and retaining a linear model should at least capture the overall increasing trend of the linewidth with frequency (Houdek et al. 1999), although it might not be even be strictly monotonic (Chaplin et al. 2005). The relevant parameter space is now  $\{\mathbf{v}, \Gamma_1, \Gamma_2, \mathbf{H}\}$ . Assuming that the  $\nu_k$ ,  $\Gamma_1$  and  $\Gamma_2$  are all independent parameters, we can write the second term in (6) as

$$\begin{aligned} \pi(\mathbf{v}, \Gamma, \mathbf{H}) &= \pi(\mathbf{H}|\mathbf{v}, \Gamma_1, \Gamma_2)\pi(\mathbf{v}, \Gamma_1, \Gamma_2) \\ &= \pi(\mathbf{H}|\mathbf{v}, \Gamma_1, \Gamma_2) \prod_{k=1}^K \pi(\nu_k) \prod_{i=1}^2 \pi(\Gamma_i). \end{aligned} \quad (7)$$

Mode heights have been fixed as follows. We first smoothed the power spectrum<sup>8</sup>, according to Kjeldsen et al. (2008). To each  $\nu_k$  we can thus associate a value for the amplitude,  $A_k$ , which is then converted to height using the relation  $H_k = 2A_k^2/(\pi\Gamma_k)$  (Chaplin et al. 2008). The first term in the right-hand side of Eq. (7) is therefore

$$\pi(\mathbf{H}|\mathbf{v}, \Gamma_1, \Gamma_2) = \prod_{k=1}^K \delta(H_k - \frac{2A_k^2}{\pi\Gamma_k}). \quad (8)$$

For the frequencies and the mode lifetimes we chose uniform distributions<sup>9</sup> on the regions where the parameters are allowed to vary. The prior density simplifies to

$$\pi(\theta) = \prod_{k=1}^K \mathbb{1}_{a_k}(\nu_k) \delta(H_k - \frac{2A_k^2}{\pi\Gamma_k}) \prod_{i=1}^2 \mathbb{1}_{b_i}(\Gamma_i), \quad (9)$$

<sup>7</sup>This burn-in sequence is then discarded when estimating the PPD.

<sup>8</sup>This procedure might, of course, depend slightly on the filter used. We neglect this contribution.

<sup>9</sup>Such distributions sometimes enter in the quite generic category of uninformative priors. Considering that we severely restrict our individual modes to vary in a small portion of the frequency domain, this seems hardly to apply here...

where

$$\mathbb{1}_a(\nu) = \begin{cases} 1 & \text{if } \nu \in a, \\ 0 & \text{if } \nu \notin a. \end{cases} \quad (10)$$

The intervals  $a_k$  were chosen to have widths of 12  $\mu\text{Hz}$  and be centred on the initial guesses. These were selected by multiplying by 0.991 the solar frequencies from Broomhall et al. (2009, see Sect. 4.3.4 for a discussion). This ensured that the daily aliases at  $\pm 11.57 \mu\text{Hz}$  of each modes should not perturb the fit. Moreover, provided that the small separations (see Sect. 5.3) are large enough, this should also limit the possibilities of “mode swapping” for close peaks. We consider frequencies only in the range 1965–3700  $\mu\text{Hz}$ . The total number of frequencies was fixed at  $K = 52$ . This prior effectively filters out the signal that is not included in windows centred on our first guesses for the frequencies. For the linewidths, we set  $b_1 = b_2 = [0, 5] \mu\text{Hz}$ .

This is essentially the same approach as used for the study of Procyon with the same algorithm in Bedding et al. (2010). One has to keep in mind that these are strong assumptions, especially the one on the mode heights. This has to be taken into consideration when comparing the results to other methodologies in Sect. 5.

The simulation made use of 6 parallel chains for tempering. After removal of a  $\sim 250\,000$  sample burn-in sequence, we obtained an approximate PPD for  $\theta$  based on  $\sim 2\,000\,000$  samples.

#### 4.3.2. Statistical analysis

The MCMC simulation provides us directly with the marginal probability densities of our parameters<sup>10</sup>. We can thus calculate various statistics that describe these densities. For each of them, we considered the posterior mode and the posterior median.

The corresponding  $100(1 - \eta)\%$  credible sets for the modes was defined as the smallest set containing the parameter with probability  $1 \geq 1 - \eta \geq 0$

$$C_k^{\text{mode}} = \{\nu_k \in a_k | \pi(\nu_k|\mathbf{y}) \geq q(\eta)\}, \quad (11)$$

with  $q$  the smallest constant such as  $P(C_k^{\text{mode}}|\mathbf{y}) \geq 1 - \eta$ . This credible set always include the mode<sup>11</sup>.

For the medians, the credible 100% sets were defined as the intervals for which the parameter is higher or lower to these quantities with equal probability,  $\beta/2$  with  $1 \geq \beta \geq 0$ . We computed them using the cumulative distribution function<sup>12</sup> (cdf)

$$C_k^{\text{median}} = \{\nu_k \in a_k | \frac{1 + \beta}{2} \geq \text{cdf}(\nu_k|\mathbf{y}) \geq \frac{1 - \beta}{2}\}. \quad (12)$$

The corresponding credible intervals are simply the upper and lower values of the corresponding set. We now discuss the relevance of these statistics using our MCMC algorithm on simulated time series.

<sup>10</sup>The marginal PPD of a given parameter is simply the PPD integrated over all the other

$$\pi(\theta_i|\mathbf{y}) = \int \pi(\theta|\mathbf{y}) d\theta_1 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_N$$

<sup>11</sup>The marginal distribution itself was estimated from the MCMC sample using a kernel density estimation algorithm.

<sup>12</sup>The cdf is obtained directly from the MCMC output

#### 4.3.3. Results for simulated time series

Evaluating the performance of estimation methods against artificial data is common practice. Of course, we need simulations realistic enough to ensure the conclusions can be transposed to the real case. 18 Sco offers a good opportunity to carry out such an exercise. Indeed, its similarity to the Sun suggests that we can obtain satisfying mock data. To that effect, we simply need to start from a solar seismic model, which, with respect to asteroseismic standards, is extremely well-known, and scale its parameters accordingly.

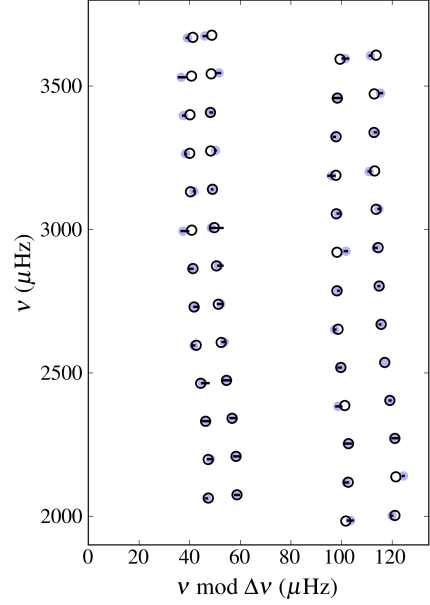
In this section, we use only one artificial time series, in particular because of the MCMC algorithm is time-consuming. The artificial time series were constructed using the BiSON solar frequencies. They were multiplied by the ratio of 18 Sco average large separation to the solar one ( $\approx 0.991$ ). The mode heights and lifetimes were assumed to be solar. These parameters were given as input to the solarFLAG simulator (Jiménez-Reyes et al. 2008) to produce evenly sampled time series. These data were subsequently interpolated to the observing times of the 18 Sco data.

An additional heuristic adjustment has been made in order to produce a realistic data set. We scaled the amplitude of the time series by a constant factor and added to each point the realization of a Gaussian random variable to simulate the photon noise. This was done by setting the noise level in the frequency space and then converting to the time space using the relation  $\sigma_{\text{phot}} = NA(\nu)/\pi$ , with  $\sigma_{\text{phot}}$  the scatter due to the photon noise,  $N$  the length of the time series, and  $A(\nu)$  the photon noise level as estimated from the amplitude spectrum. The scaling constant and the variance of the underlying noise distribution were adjusted to reproduce roughly the signal-to-noise ratio observed in the amplitude spectrum of 18 Sco<sup>13</sup>. We crudely fixed these values to 0.95 and 0.4 cm/s, so that the artificial signal-to-noise ratio is 6.46, compared to the 6.65 value measured for the observations. This gives a maximum amplitude and a noise level in the simulated spectrum within 5% of the observed ones.

We note that such a photon noise level corresponds to an average amplitude of 2.7 cm/s (or  $8.3 \times 10^{-5} \text{ m}^2 \text{ s}^{-5} / \mu\text{Hz}$ ) in the region 1800–2200  $\mu\text{H}$ . This is similar to the observed value, for which we assumed that only photon noise was present in this interval (Sect. 3). This might be an *a posteriori* warning that this assumption is inadequate. It might also mean that neglecting the term  $\mathcal{P}'(\nu)$  in Eq. (3) is not valid. The time series was analyzed using the MCMC algorithm set as described above.

Table 2 gives the results of our numerical experiment. It displays the mode and median estimates of the frequencies and their corresponding credible intervals. We also listed the input frequencies to the time series simulator. These results might guide us as to which summary statistic to use. The median capturing the central tendency of the distribution, it might often lead to a better fit to the data than the mode of the marginal posterior. However, we notice that 35% of the input frequencies are contained in the credible intervals for the modes, whereas only 30% of them are in the median credible intervals. The mode credible intervals are in general larger than those computed for the median, and quite often the former encompass the latter. These low values might indicate that we underestimate the uncertainties in both cases. On the other hand, if we define an average absolute error as  $\sum_k |\hat{\nu}_k - \nu_k^{\text{input}}| / K$  with  $\hat{\nu}_k$ ,  $k = 1, \dots, K$ , the estimated fre-

<sup>13</sup>Defined for practical purpose as the ratio of the amplitude maximum in the 1500–3700  $\mu\text{Hz}$  region to the averaged amplitude over the 1800–2200  $\mu\text{Hz}$  interval.



**Fig. 4.** Echelle diagram for 18 Sco. The blue dots represent the frequencies given in Table 3. The 68.3% credible interval are also represented. The black circles show the scaled BiSON frequencies for the Sun (their best-fitting model, Table 1 in Broomhall et al. 2009). To plot this échelle diagram, we used the value of the large separation given in Paper I.

quencies, we find a larger value for the modes (1.76  $\mu\text{Hz}$ ) than for the medians (1.31  $\mu\text{Hz}$ ). This indicates that the median is indeed, on average, more accurate than the mode.

A closer look at the marginal densities could give some explanation to this. We often observe very skewed or multimodal marginal posterior densities for the eigenfrequencies of our model. It is especially true at low and high frequency, i.e. for low and high radial orders. The median is a better estimator in ~66% of the case, almost all corresponding to this kind of difficult situations. Therefore, and considering that the credible intervals are conservative enough for both statistics, we opted for the estimator leading to the most accurate and precise values, i.e. the median.

It should also be noted that, at low frequencies, some densities are very strongly peaked around a central eigenfrequency. This translates into very narrow credible intervals that seem to be unrealistic. This can be explained by the fact that these pulsation modes have long lifetimes and are thus unresolved by our time series. Thus, they appear in the spectrum as Dirac functions and are not properly modelled by a Lorentzian. They should be considered carefully. At the very least the credible intervals associated with these particular frequencies are meaningless.

#### 4.3.4. Mode identification

The problem of identification, i.e. associating a couple  $(n, l)$  to the frequencies<sup>14</sup> detected in the time series, can be extremely difficult to solve. The standard way to proceed consists in comparing the observations to the theoretical asymptotic formula (see note 6). However, there are many instances of seismic stud-

<sup>14</sup>Rotation is neglected, therefore  $m = 0$ .



**Table 2.** Results from the benchmarking of the estimation algorithms. For the right and left tables, columns 1 and 2 give the orders and degrees of the simulated modes. Column 3 gives the input frequencies to the time series simulator (Jiménez-Reyes et al. 2008, only the frequencies we effectively searched for are displayed, more modes with lower amplitude were included in the simulated time series). Column 4 and 5 display the frequencies obtained from our MCMC algorithm using the posterior median (Col. 4) and the mode (Col. 5) estimates alongside with the upper and lower limits of the corresponding credible interval. Column 6 shows the frequencies estimated with the MAP direct-optimization methodology, alongside uncertainties derived by inverting the Hessian matrix.

$l$	$n$	Input	Median	Mode	MAP	$l$	$n$	Input	Median	Mode	MAP
0	14	2074.60	2072.33 <sup>+0.71</sup> <sub>-0.33</sub>	2072.00 <sup>+1.62</sup> <sub>-0.85</sub>	2074.39 ± 0.39	2	13	2063.21	2061.33 <sup>+2.20</sup> <sub>-0.57</sub>	2060.73 <sup>+0.20</sup> <sub>-0.20</sub>	2064.18 ± 0.43
0	15	2208.52	2207.75 <sup>+0.08</sup> <sub>-0.13</sub>	2207.78 <sup>+0.21</sup> <sub>-0.19</sub>	2208.11 ± 0.18	2	14	2197.53	2196.48 <sup>+0.09</sup> <sub>-0.18</sub>	2196.54 <sup>+0.22</sup> <sub>-0.19</sub>	2197.01 ± 0.14
0	16	2341.62	2341.60 <sup>+0.34</sup> <sub>-0.51</sub>	2341.76 <sup>+0.55</sup> <sub>-0.62</sub>	2341.98 ± 0.42	2	15	2331.10	2332.79 <sup>+0.55</sup> <sub>-1.72</sub>	2332.96 <sup>+0.99</sup> <sub>-0.25</sub>	2330.55 ± 0.34
0	17	2473.77	2474.77 <sup>+0.65</sup> <sub>-0.62</sub>	2474.77 <sup>+1.36</sup> <sub>-1.17</sub>	2473.56 ± 0.33	2	16	2463.53	2463.47 <sup>+0.52</sup> <sub>-0.61</sub>	2463.57 <sup>+1.34</sup> <sub>-1.39</sub>	2463.76 ± 0.53
0	18	2606.07	2606.27 <sup>+0.37</sup> <sub>-0.47</sub>	2606.71 <sup>+0.50</sup> <sub>-1.27</sub>	2606.10 ± 0.08	2	17	2596.15	2595.29 <sup>+0.30</sup> <sub>-0.46</sub>	2595.64 <sup>+0.55</sup> <sub>-1.18</sub>	2596.00 ± 0.08
0	19	2739.37	2738.42 <sup>+0.38</sup> <sub>-0.33</sub>	2738.22 <sup>+1.06</sup> <sub>-0.86</sub>	2738.57 ± 0.16	2	18	2729.73	2726.75 <sup>+0.29</sup> <sub>-0.37</sub>	2726.94 <sup>+0.70</sup> <sub>-0.89</sub>	2727.57 ± 0.17
0	20	2872.82	2872.56 <sup>+0.86</sup> <sub>-0.57</sub>	2871.87 <sup>+2.08</sup> <sub>-0.82</sub>	2872.38 ± 0.13	2	19	2863.47	2861.20 <sup>+0.33</sup> <sub>-0.29</sub>	2861.00 <sup>+0.92</sup> <sub>-0.65</sub>	2863.65 ± 0.15
0	21	3006.50	3002.07 <sup>+1.22</sup> <sub>-0.57</sub>	3001.11 <sup>+1.93</sup> <sub>-0.59</sub>	3005.58 ± 0.26	2	20	2997.50	2996.55 <sup>+0.40</sup> <sub>-0.48</sub>	2996.97 <sup>+0.75</sup> <sub>-1.39</sub>	2997.59 ± 0.24
0	22	3140.17	3141.63 <sup>+0.30</sup> <sub>-0.33</sub>	3141.70 <sup>+0.77</sup> <sub>-0.77</sub>	3140.86 ± 1.18	2	21	3131.46	3131.08 <sup>+0.16</sup> <sub>-0.17</sub>	3131.11 <sup>+0.39</sup> <sub>-0.42</sub>	3131.55 ± 0.39
0	23	3273.89	3275.11 <sup>+0.36</sup> <sub>-0.36</sub>	3275.05 <sup>+0.98</sup> <sub>-0.86</sub>	3274.67 ± 0.73	2	22	3265.46	3263.80 <sup>+0.35</sup> <sub>-0.41</sub>	3264.08 <sup>+0.73</sup> <sub>-1.11</sub>	3264.55 ± 0.52
0	24	3408.19	3406.99 <sup>+0.49</sup> <sub>-0.41</sub>	3406.70 <sup>+1.19</sup> <sub>-0.95</sub>	3407.34 ± 0.28	2	23	3400.03	3398.69 <sup>+0.31</sup> <sub>-0.34</sub>	3398.88 <sup>+0.72</sup> <sub>-0.89</sub>	3399.34 ± 0.46
0	25	3542.70	3542.89 <sup>+0.88</sup> <sub>-1.05</sub>	3543.37 <sup>+1.51</sup> <sub>-2.86</sub>	3540.86 ± 1.27	2	24	3534.80	3530.31 <sup>+0.65</sup> <sub>-0.49</sub>	3529.53 <sup>+1.62</sup> <sub>-0.58</sub>	3531.10 ± 2.02
0	26	3677.54	3674.49 <sup>+0.70</sup> <sub>-0.72</sub>	3674.42 <sup>+1.42</sup> <sub>-1.86</sub>	3675.50 ± 1.88	2	25	3669.86	3667.57 <sup>+1.58</sup> <sub>-1.46</sub>	3665.18 <sup>+4.32</sup> <sub>-0.83</sub>	3665.32 ± 2.00
1	13	2002.44	2001.50 <sup>+0.03</sup> <sub>-0.04</sub>	2001.51 <sup>+0.02</sup> <sub>-0.04</sub>	2001.96 ± 0.26	3	12	1983.00	1984.31 <sup>+1.11</sup> <sub>-1.02</sub>	1985.79 <sup>+1.22</sup> <sub>-3.45</sub>	1983.74 ± 0.48
1	14	2137.33	2137.00 <sup>+0.78</sup> <sub>-0.46</sub>	2136.62 <sup>+1.34</sup> <sub>-0.58</sub>	2137.33 ± 0.43	3	13	2118.42	2115.46 <sup>+1.19</sup> <sub>-0.86</sub>	2113.83 <sup>+3.00</sup> <sub>-0.88</sub>	2118.57 ± 0.38
1	15	2271.23	2271.32 <sup>+0.18</sup> <sub>-0.47</sub>	2271.52 <sup>+0.13</sup> <sub>-0.15</sub>	2271.37 ± 0.35	3	14	2252.78	2251.42 <sup>+1.66</sup> <sub>-1.40</sub>	2250.00 <sup>+0.30</sup> <sub>-1.63</sub>	2254.28 ± 0.68
1	16	2403.73	2403.16 <sup>+0.18</sup> <sub>-0.21</sub>	2403.21 <sup>+0.35</sup> <sub>-0.69</sub>	2403.89 ± 0.39	3	15	2385.91	2385.77 <sup>+0.86</sup> <sub>-0.91</sub>	2386.00 <sup>+2.59</sup> <sub>-2.27</sub>	2386.54 ± 0.82
1	17	2536.27	2540.48 <sup>+0.49</sup> <sub>-0.37</sub>	2541.21 <sup>+0.34</sup> <sub>-1.32</sub>	2536.61 ± 0.16	3	16	2518.85	2519.01 <sup>+1.00</sup> <sub>-0.87</sub>	2518.67 <sup>+3.06</sup> <sub>-1.32</sub>	2519.77 ± 0.18
1	18	2669.12	2669.39 <sup>+0.09</sup> <sub>-0.09</sub>	2669.41 <sup>+0.19</sup> <sub>-0.23</sub>	2669.92 ± 0.21	3	17	2652.01	2649.85 <sup>+0.34</sup> <sub>-0.33</sub>	2649.73 <sup>+0.89</sup> <sub>-0.91</sub>	2650.13 ± 0.22
1	19	2802.69	2802.13 <sup>+0.11</sup> <sub>-0.11</sub>	2802.14 <sup>+0.26</sup> <sub>-0.28</sub>	2803.04 ± 0.18	3	18	2786.05	2789.10 <sup>+0.53</sup> <sub>-0.49</sub>	2789.09 <sup>+1.89</sup> <sub>-0.92</sub>	2787.33 ± 0.23
1	20	2936.64	2936.45 <sup>+0.13</sup> <sub>-0.13</sub>	2936.42 <sup>+0.34</sup> <sub>-0.29</sub>	2937.02 ± 0.18	3	19	2920.37	2923.79 <sup>+0.60</sup> <sub>-0.81</sub>	2924.46 <sup>+1.10</sup> <sub>-2.03</sub>	2921.05 ± 0.47
1	21	3070.29	3072.04 <sup>+1.14</sup> <sub>-1.64</sub>	3073.84 <sup>+1.00</sup> <sub>-1.64</sub>	3074.67 ± 0.54	3	20	3054.48	3054.85 <sup>+1.28</sup> <sub>-2.45</sub>	3056.51 <sup>+0.78</sup> <sub>-0.95</sub>	3057.32 ± 0.62
1	22	3204.20	3202.73 <sup>+0.28</sup> <sub>-0.28</sub>	3202.78 <sup>+0.62</sup> <sub>-0.77</sub>	3203.02 ± 0.51	3	21	3188.85	3187.85 <sup>+1.21</sup> <sub>-1.12</sub>	3187.58 <sup>+2.59</sup> <sub>-2.87</sub>	3184.63 ± 1.26
1	23	3338.27	3338.60 <sup>+0.22</sup> <sub>-0.21</sub>	3338.49 <sup>+0.61</sup> <sub>-0.41</sub>	3339.05 ± 0.31	3	22	3323.28	3324.29 <sup>+1.31</sup> <sub>-1.25</sub>	3323.05 <sup>+4.40</sup> <sub>-0.98</sub>	3322.36 ± 0.46
1	24	3472.67	3471.80 <sup>+0.61</sup> <sub>-1.12</sub>	3472.38 <sup>+0.82</sup> <sub>-1.12</sub>	3472.51 ± 1.13	3	23	3458.15	3457.42 <sup>+1.92</sup> <sub>-1.01</sub>	3456.20 <sup>+1.35</sup> <sub>-1.30</sub>	3456.01 ± 1.20
1	25	3607.63	3604.87 <sup>+1.02</sup> <sub>-0.91</sub>	3603.29 <sup>+0.42</sup> <sub>-0.19</sub>	3605.10 ± 2.22	3	24	3593.50	3595.80 <sup>+1.09</sup> <sub>-1.34</sub>	3596.28 <sup>+3.22</sup> <sub>-2.22</sub>	3587.87 ± 2.49

ies, both from space and ground-based, whose outcome led to uncertainty on the identification (Martić et al. 2004; Carrier & Bourban 2003; Kjeldsen et al. 2005; Appourchaux et al. 2008; Benomar et al. 2009; Bedding et al. 2010). This of course may be pathological when confronted with single-site ground-based observations. This is problematic from the modelling standpoint, because if one wants to compare observed and theoretical frequencies (or, see Sect. 5.3, rational functions of the  $\nu_{n,l}$ ), one needs to be able to identify precisely the observed modes.

We are thus in a very particular position when studying 18 Sco, in which we can confirm *a posteriori* the orders and degrees we assigned for each mode. If we make the assumption that in a close vicinity of the Sun the surface related terms in the asymptotic relation do not vary significantly, we can consider that the individual frequencies scale roughly as the average large separation, i.e. as the mean density of the star.

Such a situation is expected if the stars are close to homologous (which in fact is never really the case for main-sequence stars). This is somewhat justified by the findings of Paper I, giving a ratio for the large separations  $\Delta\nu_{\odot}/\Delta\nu_{18\text{Sco}} \sim 1.0007$ . Therefore, we simply assign the  $(n, l)$  by comparing directly the frequencies of 18 Sco to the scaled solar ones as suggested by Bedding & Kjeldsen (2010). A graphical check is sufficient to

do so, as seen in Fig. 4, in which are plotted the solar modes obtained from BiSON (Broomhall et al. 2009).

#### 4.3.5. Observed frequencies

We analyzed the observed time series using the method described in Sect. 4.3.1. For the 52 fitted modes, we found credible intervals consistent with those estimated from the simulated time series. Given the comments made above, we are confident in their robustness. In some cases we obtained very narrow credible sets, whose upper and lower limits, relative to the median value, are  $< 0.1 \mu\text{Hz}$  in absolute value. These are always high amplitude modes. In Table 3, we give the medians of the marginal densities alongside the corresponding 68.3% credible regions. They are shown in Fig. 4.

The marginal probability densities are qualitatively in good agreement with the results from the simulations. However more of them exhibit multiple maxima, perhaps because of the lower photon noise level used in the simulations. This is always problematic when using the median, since it might bias the estimation towards maxima not corresponding to the real frequency. We flagged these modes in Table 3. Note, however, that the process remains subjective, and we considered only the densities for

**Table 3.** Oscillation frequencies detected for 18 Sco along with their radial order  $n$  and degree  $l$ . The quantities given here are the median of the marginal posterior distributions for each eigenfrequency considered in our model. The corresponding credible intervals are given alongside. We use some flags to signal estimates requiring additional caution. (\*) indicate eigenfrequencies with clear multimodal marginal distributions, (†) pulsation modes potentially unresolved, (‡) pulsation modes that have not passed a hypothesis testing when using the heights estimated with the MAP approach in Sect. 5.1.2 and (?) the mode that has not been identified by the MAP approach.

$n$	$l = 0$	$l = 1$	$l = 2$	$l = 3$
12				1985.12 <sup>+0.87</sup> <sub>-1.12</sub> (‡)
13		2001.51 <sup>+0.02</sup> <sub>-0.02</sub> (†)	2062.78 <sup>+0.03</sup> <sub>-0.20</sub> (†)	2117.89 <sup>+0.60</sup> <sub>-0.99</sub> (*)
14	2074.41 <sup>+1.05</sup> <sub>-0.40</sub> (†)	2140.38 <sup>+0.03</sup> <sub>-0.02</sub> (†)	2198.63 <sup>+0.88</sup> <sub>-0.92</sub> (*)	2253.00 <sup>+1.02</sup> <sub>-1.64</sub>
15	2208.79 <sup>+1.44</sup> <sub>-0.77</sub> (*, ‡)	2271.66 <sup>+1.12</sup> <sub>-1.18</sub> (*)	2331.20 <sup>+1.57</sup> <sub>-1.02</sub> (*)	2383.34 <sup>+1.35</sup> <sub>-0.68</sub> (*, ‡)
16	2342.00 <sup>+0.57</sup> <sub>-1.10</sub>	2403.96 <sup>+0.07</sup> <sub>-0.20</sub> (*)	2464.27 <sup>+2.40</sup> <sub>-0.17</sub> (*)	2518.43 <sup>+0.49</sup> <sub>-0.42</sub>
17	2473.63 <sup>+1.79</sup> <sub>-1.00</sub> (*)	2536.58 <sup>+0.04</sup> <sub>-0.04</sub>	2595.27 <sup>+0.26</sup> <sub>-0.31</sub> (*, ‡)	2650.84 <sup>+0.35</sup> <sub>-0.36</sub>
18	2607.31 <sup>+0.26</sup> <sub>-0.75</sub> (‡)	2668.72 <sup>+0.11</sup> <sub>-0.14</sub>	2730.13 <sup>+0.74</sup> <sub>-0.24</sub>	2786.75 <sup>+0.07</sup> <sub>-0.08</sub>
19	2740.27 <sup>+0.31</sup> <sub>-1.01</sub> (*)	2802.67 <sup>+0.09</sup> <sub>-0.08</sub>	2863.20 <sup>+0.27</sup> <sub>-0.73</sub> (*)	2924.12 <sup>+0.31</sup> <sub>-0.48</sub>
20	2873.78 <sup>+1.61</sup> <sub>-0.12</sub> (*)	2936.02 <sup>+0.22</sup> <sub>-0.29</sub>	2994.18 <sup>+2.01</sup> <sub>-0.48</sub> (*)	3055.39 <sup>+0.17</sup> <sub>-0.17</sub> (‡)
21	3005.34 <sup>+4.41</sup> <sub>-0.41</sub> (*, ‡)	3071.49 <sup>+0.12</sup> <sub>-0.11</sub>	3132.89 <sup>+0.07</sup> <sub>-0.07</sub>	3187.21 <sup>+1.23</sup> <sub>-1.24</sub> (‡)
22	3140.10 <sup>+0.06</sup> <sub>-0.06</sub>	3202.06 <sup>+0.53</sup> <sub>-0.19</sub> (*)	3263.98 <sup>+0.16</sup> <sub>-0.17</sub>	3322.77 <sup>+0.14</sup> <sub>-0.22</sub>
23	3275.80 <sup>+0.14</sup> <sub>-0.17</sub> (*)	3338.91 <sup>+0.13</sup> <sub>-0.14</sub>	3397.52 <sup>+0.29</sup> <sub>-0.17</sub>	3458.25 <sup>+1.68</sup> <sub>-1.75</sub> (‡)
24	3408.53 <sup>+0.36</sup> <sub>-0.26</sub> (*)	3475.34 <sup>+0.53</sup> <sub>-0.53</sub> (*)	3531.08 <sup>+1.69</sup> <sub>-0.96</sub> (?)	3595.93 <sup>+1.00</sup> <sub>-1.01</sub> (‡)
25	3545.96 <sup>+0.44</sup> <sub>-1.01</sub> (*, ‡)	3605.76 <sup>+0.40</sup> <sub>-0.27</sub>	3667.94 <sup>+0.26</sup> <sub>-0.33</sub>	
26	3674.80 <sup>+0.53</sup> <sub>-0.45</sub>			

which the secondary maxima were obvious. These modes can be used in subsequent studies, but one should keep in mind that the summary statistics we used do not capture all the features of the distribution. The average width of the 68.3% credible regions on the median is slightly smaller than the one observed for the simulated time series, but this should not prove significant. Compared to the simulations, we also found more skewed distributions. This reinforces our choice to use the median for the statistical summary.

We again observe some very sharply peaked marginal densities. Such shapes lead to very narrow and unrealistic credible regions, for instance much narrower than has been found from space missions with much longer observing baselines (see e.g. Gaulme et al. 2009; Mathur et al. 2010; Campante et al. 2011). They appear at low frequencies where the mode lifetimes are known to be longer, and we can thus express doubts about the resolution of modes (0,14), (1,13), (1,14), and (2,13), which are flagged accordingly in Table 3. They are all located at low frequencies,  $\nu \lesssim 2150 \mu\text{Hz}$ . This give an approximate limit above which the modes start to be resolved. In the case of modes (1,13) and (1,14), the kernel estimations of the densities return a numerical error. This might indeed confirm that they cannot be approximated by a continuous density, which would be characteristic of an unresolved pulsation mode. These modes might very well be real. However, because the Lorentzian model becomes incorrect for unresolved modes, the associated uncertainties are likely to be widely underestimated.

Finally, and anticipating the discussion below, some modes, when estimated with an alternative strategy which does not fix the heights (see Sect. 5) did not pass an hypothesis testing. We flagged them and do not recommend their use in subsequent studies.

The results of our MCMC simulation are available at [www.astro.up.pt/~bazot/data/18ScoII/](http://www.astro.up.pt/~bazot/data/18ScoII/).

## 5. Discussion

### 5.1. Comparison to a MAP approach

Our goal in this section is to compare the performance of our MCMC approach with another Bayesian strategy based on the direct optimization of the PPD. Note that this is a very general comparison, since not only do we change the *a posteriori* estimators (median and maximum of the PPD), but we also modify our probabilistic model, i.e. the priors. We chiefly want to get an idea of how consistent they might be. This is good procedure to cross-check results obtained using different methodologies. In the case of 18 Sco, given the relatively difficult nature of the data, we see this as necessary.

In the Maximum A Posteriori (MAP) approach, the likelihood is replaced by the PPD. The estimator for the model parameters becomes

$$\text{MAP}(\theta) = \underset{\theta}{\operatorname{argmax}}(\pi(\theta|\mathbf{y})). \quad (13)$$

This estimator is sometimes called the regularized likelihood. This is a common strategy and has been used in the case of the Sun (Chaplin et al. 2002; Broomhall et al. 2009) and stars observed from satellites (Appourchaux et al. 2008; Deheuvels & Michel 2010). Here, the regularized likelihood is maximized using a Powell algorithm.

#### 5.1.1. Setup and tests

An interesting feature of this direct optimization approach is that it converges somewhat more easily than our MCMC algorithm. Therefore, we were able to use less constraining priors. The mode lifetimes and heights were both left free to vary. We applied on the frequencies priors close to those described by Eq. (9). The main difference is that, instead of considering individual frequencies, we considered pairs of frequencies of similar parities and differing by one radial order. The  $a_k$  were set to



**Table 4.** Estimated frequencies of 18 Sco from the MAP method using direct optimization. The associated uncertainties were derived from the inversion of the Hessian matrix.

$n$	$l = 0$	$l = 1$	$l = 2$	$l = 3$
12				$1984.32 \pm 0.68$
13		$2002.00 \pm 0.26$	$2063.13 \pm 0.16$	$2118.01 \pm 0.10$
14	$2074.88 \pm 0.17$	$2140.60 \pm 0.12$	$2199.47 \pm 0.39$	$2254.95 \pm 0.24$
15	$2209.49 \pm 0.19$	$2270.54 \pm 0.17$	$2330.74 \pm 0.20$	$2385.35 \pm 2.17$
16	$2339.14 \pm 0.18$	$2404.25 \pm 0.49$	$2467.29 \pm 0.78$	$2519.03 \pm 0.44$
17	$2475.62 \pm 0.67$	$2536.84 \pm 0.11$	$2595.50 \pm 0.15$	$2651.51 \pm 1.24$
18	$2605.46 \pm 0.15$	$2669.69 \pm 0.97$	$2730.41 \pm 1.06$	$2786.96 \pm 0.20$
19	$2741.34 \pm 1.54$	$2803.55 \pm 0.11$	$2863.74 \pm 0.17$	$2918.07 \pm 1.40$
20	$2874.13 \pm 0.20$	$2936.92 \pm 0.69$	$2994.29 \pm 1.18$	$3055.77 \pm 0.61$
21	$3006.00 \pm 1.35$	$3071.97 \pm 0.48$	$3133.10 \pm 0.43$	$3186.07 \pm 1.53$
22	$3140.58 \pm 0.40$	$3203.11 \pm 0.95$	$3264.55 \pm 0.57$	$3322.89 \pm 0.40$
23	$3276.19 \pm 0.68$	$3339.44 \pm 0.37$	$3398.24 \pm 0.52$	$3459.01 \pm 2.76$
24	$3408.74 \pm 0.60$	$3476.24 \pm 1.81$	$3545.86 \pm 3.58$	$3593.56 \pm 2.68$
25	$3556.15 \pm 3.92$	$3610.70 \pm 2.32$	$3669.81 \pm 2.20$	
26	$3679.63 \pm 2.51$			

$\pm 22 \mu\text{Hz}$  above the  $l = 0$  (or 1) and below the  $l = 2$  (or 3) modes (Fletcher et al. 2009).

It should be noted that the uncertainties in the MAP framework are estimated by inverting the Hessian matrix of the parameters. This is well-justified if the errors, and their second derivatives with respect to the parameters, are small and if these derivatives are not much correlated with the errors. Moreover, the posed fitting problem must also be well-constrained, otherwise the formal uncertainties will also be a poor representation of the true uncertainties. In our case, given the relative complexity of the probability distributions we consider (as can be seen from the MCMC results), these assumptions are not likely to hold. We also note that, from one mode to the other, the Hessian-derived and MCMC uncertainties might differ significantly from the MCMC estimates.

Another issue with the MAP estimation approach is the potential instability of the minimization algorithm with respect to the initial guesses. To test this, we varied randomly the first-guess frequencies – using a top-hat distribution of width  $\pm 3 \mu\text{Hz}$  – and the final frequencies were median estimates over 1000 such fits. We found that the scatter in the estimate over these 1000 fits is similar to or larger than the Hessian-derived uncertainties. It was the possibility of deriving more realistic uncertainties that ultimately led us to choose the MCMC estimates as a reference.

We tested our MAP algorithm using the artificial data described in Sect. 4.3.3. The results are given in Table 2, alongside those from the MCMC approach. The estimated frequencies are extremely close for the two methods. Nevertheless, the MAP algorithm performs better in terms of accuracy, with the average absolute error being lower than from the MCMC estimates values.

This opens the door to questions with respect to the proper use of priors in Bayesian analysis. Indeed, if this relative lack of accuracy in the MCMC approach is caused by a bias introduced by the stringent prior constraints imposed, it means that those included in our MAP setup are better. However, this is a difficult problem, numerically, to assume such a great variation of the mode lifetimes and heights in the MCMC approach. This shows how delicate it is to choose between different Bayesian methodologies. Note however that for sufficiently long and precise measurements the two approaches should converge.

### 5.1.2. Estimates from the observations

Applying the MAP algorithm to the real data, we found results, given in Table 4 for reference, similar to those of the MCMC approach for the frequency estimates. The uncertainties are consistent with our tests using simulated time series. There is only one inconsistency in the identification between the two approaches. The ( $l = 2, n = 24$ ) mode was identified as ( $l = 0, n = 25$ ) in the MCMC framework. It is likely that the ( $l = 0, n = 25$ ) frequency returned from optimization is in fact an alias of the real mode. However, these are very low-amplitude peaks, rejected in the hypothesis testing (see below).

It should also be noted that the greater “flexibility” we have in terms of convergence has allowed us to estimate the mode heights and the lifetimes. This allowed us to perform *a posteriori* some hypothesis testing for each mode (Appourchaux et al. 2009). It is somewhat more satisfying to carry out such tests on genuinely estimated heights rather than on fixed heights derived from a filtered spectrum, such as the ones used for our MCMC simulations. Therefore, in Table 3, we flag the values with positive hypothesis testing results. This way, one can choose or not to include them when using the list of frequencies. Note that only 40 frequencies were incompatible with the  $H_0$  (null) hypothesis, which tests here the hypothesis that the peak is due to the noise in the data.

### 5.2. Comparison with time-domain modelling

Another approach commonly used in asteroseismology involves representing the signal in the time domain

$$\tilde{y}(t_n) = \sum_{k=1}^M c_k \sin(2\pi\nu_k t_n) + d_k \cos(2\pi\nu_k t_n) + \epsilon_n, \quad (14)$$

where  $(c_k, d_k)$  and  $\nu_k$  are the amplitudes and frequencies of the pulsation modes and  $\epsilon_n$ , the noise.

Comparing with model (3), some shortcomings of model (14) are clear. In particular, it does not take into account the fact that the modes have finite lifetimes. This may lead to an over-fitting of the signal in the vicinity of some modes, i.e. several sine functions being required to reproduce what is actually the dual of a Lorentzian. However, this effect clearly depends on the ratio of the characteristic damping time to the length of the time series. If it is large, then the chances are high that the

mode will be unresolved, the power excess largely being confined to one frequency bin (see Sect. 5). In this case, model (14) will be accurate enough. The unresolved-mode assumption has often been made in the case of ground-based seismic observations (Kjeldsen et al. 1995; Bouchy & Carrier 2001; Bouchy et al. 2005; Bazot et al. 2007; Kjeldsen et al. 2005; Bedding et al. 2010). Because of its simplicity, interesting methodologies can be applied to the problem of estimating its parameters. We consider two here. The first one is the well-known CLEAN algorithm (Gray & Desikachary 1973; Roberts et al. 1987), used for iterative deconvolution, and the second one is the SparSpec algorithm, a penalization approach to minimization in the context of spectral analysis (Bourguignon et al. 2007). We used both methods to make sure that the results discussed below are not due to algorithmic artifact.

The objective of this section is to compare our results with simulations in order to understand the impact of our choice for a physical model (time-domain representation against frequency domain representation) for the power spectrum. We also try to understand, at least qualitatively if this is more important than our choice for the priors on the parameters included in our probabilistic description. We explain how, in the case of 18 Sco, a frequency-domain representation is an improvement for gapped and irregularly sampled time series, for which models such as Eq. (14) have previously been used.

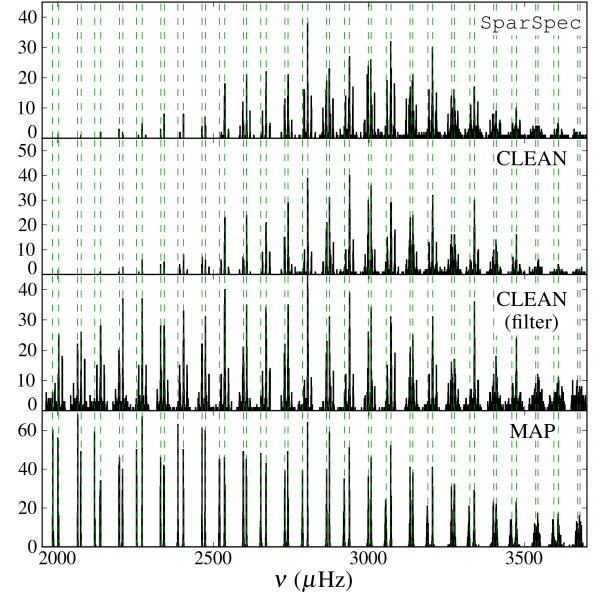
### 5.2.1. Performances of the methods

We used a sample of 100 artificial time series constructed as described in Sect. 4.3.3. They only differ by the realizations of the low-frequency and white noises. We applied the MAP, CLEAN and SparSpec algorithms to each element of this sample. The MAP setup is similar to the one described in Sect. 5.1.1. In the case of CLEAN, we limited our search to the 1500–3700  $\mu\text{Hz}$  region. We set a threshold for the detection at three times the noise level in the 1800–2200  $\mu\text{Hz}$  interval, that is  $7.2 \text{ cm.s}^{-1}$ . The relevant parameter for SparSpec is the penalization factor (Bourguignon et al. 2007), which we empirically set to 0.34, so that the results are close to those obtained for CLEAN.

Note that we did not include the MCMC algorithm in this comparison. This should not be a problem since we have seen that the results from the direct MAP optimization and the MCMC agree well. Our main goal is to evaluate the efficiency of our algorithm, i.e. how the estimated frequencies reproduce the input frequencies to the time series simulator. We are not concerned with the uncertainties on the parameters here, which were the main reason to retain the MCMC estimates as our reference. Therefore, since we have seen in Sect. 5.1 that the MAP and MCMC approach lead to close enough estimates, one can extrapolate the following discussion to the MCMC case.

In the case of CLEAN and SparSpec, it is not possible to go through the normal adjustment of the outputs of the algorithms for the 100 realizations of the time series<sup>15</sup>. We can only obtain crude estimates of how many times each input frequency is actually detected. This can be done by looking at Fig. 5, which represents the histograms of the 100 outputs for each method. These are upper limits to the rate of detection of the time-series simulator input frequencies, mostly because of the multiple peaks sometimes necessary to describe a single mode.

<sup>15</sup>These adjustments consist in removing manually the peaks that are obviously due to noise or aliases not properly removed by the algorithm. In a sense, this is very much similar to applying some *prior* knowledge one would have on the frequencies, but after the estimation process.



**Fig. 5.** Histograms of the output frequencies from the MAP, CLEAN (with and without prior filtering) and SparSpec algorithms for 100 artificial time series. The vertical green lines mark the input frequencies to the time-series simulator.

The two methods lead to very similar results. This indicates that the model we used is the main factor determining the outcome of the estimation process. Both methodologies are obviously very sensitive to the amplitude of the mode. Only in the 2800–3400  $\mu\text{Hz}$  region do these algorithms find the input frequency  $\sim 50\%$  of the time within a frequency interval corresponding to twice the natural resolution  $\delta = 1/T$ . This percentage, drops strongly below and above these values.

In the MAP case, the main factor affecting the efficiency of the algorithm is the mode lifetime. More precisely, at higher frequencies, when the mode is most likely resolved, the proportion of correct detections decreases. It is in the range 60%–100% for frequencies between 1980  $\mu\text{Hz}$  and  $\sim 3400 \mu\text{Hz}$ . It drops significantly for eigenfrequencies above  $\sim 3400 \mu\text{Hz}$ . In any case, the detection rates are much higher than those observed for CLEAN and SparSpec. This is the reflection of the fact that we used different physical models to describe the signal. It is somewhat in disagreement with the findings of White et al. (2010).

### 5.2.2. Impact of the prior formulation

To further understand how the priors affect the results, we tried to combined our CLEAN algorithm with constraints similar to those described by our prior on the frequencies. It should be noted that algorithms such as CLEAN have not been designed with a Bayesian perspective in mind (see for instance Schniter et al. 2009). Therefore, we could only try to mimic the impact of the prior. To this effect, we retained the idea that the frequency prior acts in analogy like a bandpass filter, which removes all signal outside the top-hat functions. We therefore applied such a filter (a sum of bandpass filters) to our spectrum. We then used the CLEAN algorithm to search only for two frequencies per individual bandpass filter. In a sense, this strategy is very similar to the “ridge search” approach used for  $\eta$  Boo by Kjeldsen

et al. (1995). The corresponding output histogram is displayed in Figure 5. The information we get from this test is, of course, only qualitative, but it gives an interesting picture of the performance of the two algorithms under constraints that are fairly similar.

We performed this test on the same 100 time-series sample. We can see that this definitely enhances the performance of the CLEAN algorithm. The detection rate increases everywhere, particularly in the low-frequency regions of the spectrum. In the high-frequency regions, the situation also improves, but the model is subject to limitations concerning the mode life-times, which perturb the estimation. However, the overall performance remains largely inferior to the outcome of the MAP strategy, which uses model (3). This is very revealing as to the effect the priors have on the final frequency estimates. It is often contended that Bayesian analysis may use too strong priors and retrieve only what has been defined in  $\pi(\theta)$  before the estimation. This is not entirely the case here. The priors on the frequencies we used are not so strong that any algorithm will be able to perform equally well under such a constraint. This result illustrates the subtle interplay between the numerical and statistical advantages of the Bayesian method mentioned in Sect. 4.2. Not only do the priors tighten the relevant volume in the space of parameters, but they also stabilize the fit to the data when using a more complex but also more accurate model, involving a larger number of parameters (higher dimension of the parameter space).

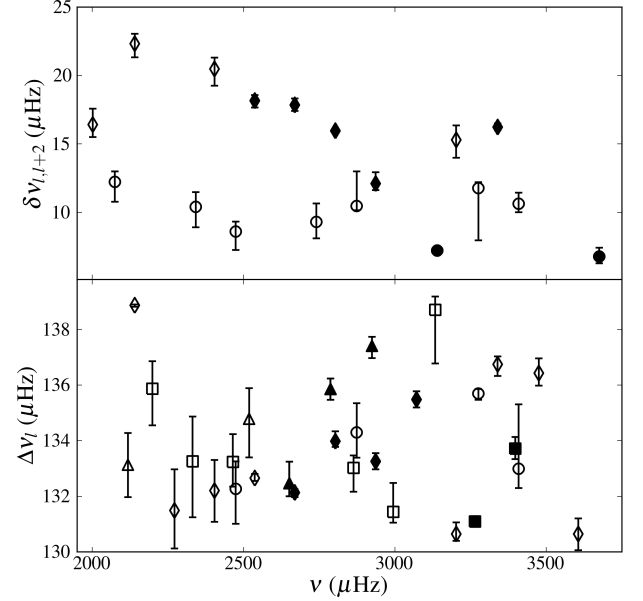
This is not the first time that Bayesian methods have been used on time series with such short time baseline (Brewer et al. 2007, for instance). However, the very favourable case of 18 Sco allows us to contend that, provided the frequency priors are accurate enough, direct fits to the power spectrum are more efficient than classical time-domain modelling. A further step would be to test this claim with more sophisticated models for the spectrum (Stahn & Gizon 2008) or the time series (Brewer & Stello 2009).

### 5.3. Large and small separations

Two common seismic indicators are the large and small separations, defined respectively by  $\Delta\nu_l(n) = \nu_{n+1,l} - \nu_{n,l}$  and  $\delta\nu_{l,l+2}(n) = \nu_{n,l} - \nu_{n-1,l+2}$ . They both stem from a first analysis of the asymptotic relation for p modes. Their use has been popularized by the fact that they are supposed to be relatively free of the unknown surface effects affecting the oscillation frequencies. This however is only partially true and other combination of frequencies have been suggested in the literature (e.g., Roxburgh & Vorontsov 2003; Cunha & Metcalfe 2007). We nevertheless limit ourselves to these two quantities, which are plotted in Fig. 6.

It is also interesting to compute the average values of the large and small separations. In order to estimate them, we use the MCMC samples. They are convenient to study densities of averages because, if we assume that the central-limit theorem roughly applies, we can expect to deal with Gaussian distribution (Benomar, private communication). This greatly simplifies the subsequent statistical analysis.

To estimate the average large separation we retained only the unflagged modes in Table 3. Basing ourselves on the asymptotic relation, we consider that, for a fixed degree, the frequency is a linear function of the mode order, with slope the average large separation. We thus computed the derivative of  $\nu(n)$  at each order for each degree and averaged over both quantities. The resulting distribution is well approximated by a Gaussian and we obtained  $\langle\Delta\nu_{0,2}\rangle = 133.8 \pm 0.2 \mu\text{Hz}$ . This value agrees with the estimate of Paper I,  $\langle\Delta\nu_{0,2}\rangle = 134.4 \pm 0.3 \mu\text{Hz}$ , at  $2\sigma$  level.



**Fig. 6.** Individual large (lower panel) and small (upper panel) separations for 18 Sco. The circles ( $\circ$ ) corresponds to  $l = 0$ , the diamonds ( $\diamond$ ) to  $l = 1$ , the squares ( $\square$ ) to  $l = 2$  and the triangles ( $\triangle$ ) to  $l = 3$ . Filled symbols mark the combinations using unflagged frequencies in Table 3. A few error bars do not appear because they are smaller than their respective symbols.

This sheds a new light on the mass estimate we gave in Paper I. Using these new values for the average large separation and the interferometric radius, one might evaluate the mass of 18 Sco to be  $1.01 \pm 0.03 M_{\odot}$ , which brings it even “closer” to the Sun. Although this agrees within the  $1\sigma$  error bars with the value of Paper I, it is relevant for modelling if one is to use on of these masses estimates in order to, for instance, apply a prior on the mass when modelling 18 Sco in the Bayesian framework (Bazot et al. 2008).

The average small separations are slightly more problematic. This is due to the fact that fewer values are available to average over, in particular for  $\delta\nu_{0,2}$ . We thus decided to include the frequencies from Table 3 flagged with a (\*). Even though they display the multiple maxima, our idea is that the most prominent peaks in the distributions will be the main contributors to the density of the averaged value. We indeed found Gaussian distributions for  $\langle\delta\nu_{0,2}\rangle$  and  $\langle\delta\nu_{1,3}\rangle$  (the latter being much better approximated by such a distribution than the former). Using their first moments, we get  $\langle\delta\nu_{0,2}\rangle = 9.4 \pm 0.9 \mu\text{Hz}$  and  $\langle\delta\nu_{1,3}\rangle = 16.7 \pm 0.8 \mu\text{Hz}$ .

It is known that the small separations depend on frequency. However, to the first order, they can be approximated by constants  $\delta\nu_{0,2}(n) \approx 6D_0 \approx 3\delta\nu_{1,3}/5$ , with  $D_0$  an integral containing the derivative of the sound speed (Gough 1986; Gabriel 1989). The estimated values lead to a ratio  $\langle\delta\nu_{0,2}\rangle/\langle\delta\nu_{1,3}\rangle = 0.57$ , in good agreement with the theoretical expectations.

## 6. Conclusion

We presented a detailed analysis of the ground-based seismic data obtained for the solar twin 18 Sco from the high-precision spectrograph HARPS. The sampling of the time series causes se-

rious problems for stellar eigenfrequency estimation. We chose to use an MCMC algorithm in order to estimate the frequencies of 52 stellar pulsation modes. A careful examination of the posterior probability densities for each of them show that at least 21 are reliable and at least 19 others are worth consideration, even though the corresponding marginal PPDs are more difficult to analyze. 11 were rejected after comparison with a the MAP direct optimization methodology. We were able to estimate Bayesian credible intervals for these modes which reflect with some robustness the uncertainties of our data.

By comparing with other estimation methods, we have discussed how reliable are the priors we used for the estimation. On the one hand, they are constraining enough to allow us to use a (relatively) realistic model. On the other hand, they are not so restrictive that they would impede a proper estimation. We note that our methodology can be further improved by increasing the efficiency of our MCMC algorithm (which would allow to relax further the priors on the parameters) and/or by using even more accurate models (which may require more conservative priors).

The individual eigenfrequencies obtained for 18 Sco allowed us to study some basic seismic estimators, including the large separations, whose estimation of the average value was addressed earlier in Paper I. The two values agree at  $2\sigma$  level, with the new one being lower. We derived a new value for the mass of the star slightly lower than the previous one. It remains to see how much this might affect the modelling of the star.

The sampling issues of our data certainly played an important part in producing the differences observed between the various methods used in our study. A next step would be to observe this star with more than one telescope, even though its magnitude makes such a task challenging for most of the ground-based instruments now available. Furthermore, we noticed that the length of the time series might imply that we only resolve a fraction of the detected modes. This could be resolved with a longer time basis.

**Acknowledgements.** The authors would like to thank the referee for his very careful report and careful comments. We feel that he helped to improve significantly this paper. This work was co-supported by grants SFRH/BPD/47994/2008, SFRH/BD/36240/2007 and PTDC/CTE-AST/098754/2008 from FCT/MCTES and FEDER, Portugal. MB thanks O. Benomar for very interesting and fruitful discussions. A-MB and WJC acknowledge the support of the UK Science and Technology Facilities Council (STFC). Funding for the Stellar Astrophysics Centre is provided by The Danish National Research Foundation. The research is supported by the ASTERISK project (ASTERoseismic Investigations with SONG and Kepler) funded by the European Research Council (Grant agreement no.: 267864).

## References

- Anderson, E. R., Duvall, Jr., T. L., & Jefferies, S. M. 1990, *ApJ*, 364, 699
- Appourchaux, T., Michel, E., Auvergne, M., et al. 2008, *A&A*, 488, 705
- Appourchaux, T., Samadi, R., & Dupret, M. 2009, *A&A*, 506, 1
- Bazot, M., Bouchy, F., Kjeldsen, H., et al. 2007, *A&A*, 470, 295
- Bazot, M., Bourguignon, S., & Christensen-Dalsgaard, J. 2008, *Memorie della Societa Astronomica Italiana*, 79, 660
- Bazot, M., Ireland, M. J., Huber, D., et al. 2011, *A&A*, 526, L4
- Bazot, M., Vauclair, S., Bouchy, F., & Santos, N. C. 2005, *A&A*, 440, 615
- Bedding, T. R. & Kjeldsen, H. 2010, *Communications in Asteroseismology*, 161, 3
- Bedding, T. R., Kjeldsen, H., Campante, T. L., et al. 2010, *ApJ*, 713, 935
- Benomar, O., Appourchaux, T., & Baudin, F. 2009, *A&A*, 506, 15
- Bouchy, F., Bazot, M., Santos, N. C., Vauclair, S., & Sosnowska, D. 2005, *A&A*, 440, 609
- Bouchy, F. & Carrier, F. 2001, *A&A*, 374, L5
- Bourguignon, S., Carfantan, H., & Böhm, T. 2007, *A&A*, 462, 379
- Brandão, I. M., Doğan, G., Christensen-Dalsgaard, J., et al. 2011, *A&A*, 527, A37+
- Bretthorst, G. L. 2000, in *Bulletin of the American Astronomical Society*, Vol. 32, *Bulletin of the American Astronomical Society*, 1438
- Brewer, B. J., Bedding, T. R., Kjeldsen, H., & Stello, D. 2007, *ApJ*, 654, 551
- Brewer, B. J. & Stello, D. 2009, *MNRAS*, 395, 2226
- Broomhall, A., Chaplin, W. J., Davies, G. R., et al. 2009, *MNRAS*, 396, L100
- Campante, T. L., Handberg, R., Mathur, S., et al. 2011, *A&A*, 534, A6
- Carrier, F. & Bourban, G. 2003, *A&A*, 406, L23
- Cayrel de Strobel, G., Knowles, N., Hernandez, G., & Bentolila, C. 1981, *A&A*, 94, 1
- Chaplin, W. J., Elsworth, Y., Isaak, G. R., et al. 2002, *MNRAS*, 336, 979
- Chaplin, W. J., Houdek, G., Appourchaux, T., et al. 2008, *A&A*, 485, 813
- Chaplin, W. J., Houdek, G., Elsworth, Y., et al. 2005, *MNRAS*, 360, 859
- Cunha, M. S. & Metcalfe, T. S. 2007, *ApJ*, 666, 413
- De Rosa, M., Duvall, Jr., T. L., & Toomre, J. 2000, *Sol. Phys.*, 192, 351
- Deeming, T. J. 1975, *Ap&SS*, 36, 137
- Deheuvels, S. & Michel, E. 2010, *Ap&SS*, 328, 259
- Doğan, G., Brandão, I. M., Bedding, T. R., et al. 2010, *Ap&SS*, 328, 101
- Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. J. P. F. G. 2011, *A&A*, 525, A140
- Fletcher, S. T., Chaplin, W. J., Elsworth, Y., & New, R. 2009, *ApJ*, 694, 144
- Foglizzo, T. 1998, *A&A*, 339, 261
- Foglizzo, T., Garcia, R. A., Boumier, P., et al. 1998, *A&A*, 330, 341
- Gabriel, M. 1989, *A&A*, 226, 278
- Gaulme, P., Appourchaux, T., & Boumier, P. 2009, *A&A*, 506, 7
- Gough, D. O. 1986, in *Hydrodynamic and Magnetodynamic Problems in the Sun and Stars*, ed. Y. Osaki, 117
- Gray, D. F. & Desikachary, K. 1973, *ApJ*, 181, 523
- Gregory, P. C. 2005, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support* (Cambridge University Press)
- Guenther, D. B., Kallinger, T., Gruberbauer, M., et al. 2008, *ApJ*, 687, 1448
- Gustafsson, B. 1998, *Space Science Reviews*, 85, 419
- Gustafsson, B. 2008, *Physica Scripta Volume T*, 130, 014036
- Handberg, R. & Campante, T. L. 2011, *A&A*, 527, A56
- Harvey, J. 1985, in *ESA Special Publication*, Vol. 235, *Future Missions in Solar Heliospheric & Space Plasma Physics*, ed. E. Rolfe & B. Battrock, 199–208
- Hastings, W. K. 1970, *Biometrika*, 57, 97
- Houdek, G., Balmforth, N. J., Christensen-Dalsgaard, J., & Gough, D. O. 1999, *A&A*, 351, 582
- Jiménez-Reyes, S. J., Chaplin, W. J., García, R. A., et al. 2008, *MNRAS*, 389, 1780
- Kjeldsen, H., Bedding, T. R., Arentoft, T., et al. 2008, *ApJ*, 682, 1370
- Kjeldsen, H., Bedding, T. R., Butler, R. P., et al. 2005, *ApJ*, 635, 1281
- Kjeldsen, H., Bedding, T. R., Viskum, M., & Frandsen, S. 1995, *AJ*, 109, 1313
- Lefebvre, S., García, R. A., Jiménez-Reyes, S. J., Turck-Chièze, S., & Mathur, S. 2008, *A&A*, 490, 1143
- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- Martić, M., Lebrun, J., Appourchaux, T., & Korzenik, S. G. 2004, *A&A*, 418, 295
- Mathur, S., García, R. A., Catala, C., et al. 2010, *A&A*, 518, A53
- Meléndez, J. & Ramírez, I. 2007, *ApJ*, 669, L89
- Meléndez, J., Schuster, W. J., Silva, J. S., et al. 2010, *A&A*, 522, A98
- Metcalfe, T. S., Monteiro, M. J. P. F. G., Thompson, M. J., et al. 2010, *ApJ*, 723, 1583
- Metropolis, N. 1953, *J. Chem. Phys.*, 21, 1087
- Miglio, A. & Montalbán, J. 2005, *A&A*, 441, 615
- Nordlund, Å., Stein, R. F., & Asplund, M. 2009, *Living Reviews in Solar Physics*, 6, 2
- Petit, P., Dintrans, B., Solanki, S. K., et al. 2008, *MNRAS*, 388, 80
- Robert, C. P. 2007, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (Springer Texts in Statistics), 2nd edn. (Springer Verlag, New York)
- Robert, C. P. & Casella, G. 1999, *Monte Carlo Statistical Methods*, 1st edn. (Springer-Verlag)
- Roberts, D. H., Lehar, J., & Dreher, J. W. 1987, *AJ*, 93, 968
- Roxburgh, I. W. & Vorontsov, S. V. 2003, *A&A*, 411, 215
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Schniter, P., Potter, L. C., & Ziniel, J. 2009, *Fast Bayesian Matching Pursuit: Model Uncertainty and Parameter Estimation for Sparse Linear Models*
- Shine, R. A., Simon, G. W., & Hurlburt, N. E. 2000, *Sol. Phys.*, 193, 313
- Stahn, T. & Gizon, L. 2008, *Sol. Phys.*, 251, 31
- Tassoul, M. 1980, *ApJS*, 43, 469
- Title, A. M., Tarbell, T. D., Topka, K. P., et al. 1989, *ApJ*, 336, 475
- Vandakurov, Y. V. 1967, *AZh*, 44, 786
- White, T. R., Brewer, B. J., Bedding, T. R., Stello, D., & Kjeldsen, H. 2010, *Communications in Asteroseismology*, 161, 39
- Zechmeister, M. & Kürster, M. 2009, *A&A*, 496, 577